CENTRE
DE RECHERCHE
SUR LES TRANSPORTS
(CRT)
CENTRE
FOR RESEARCH
ON TRANSPORTATION

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE

C.P. 6128, SUCCURSALE A
MONTRÉAL (QUÉBEC)
CANADA
H3C 3J7

TÉLÉPHONE : (514) 343-7575
TÉLEX : 05-24146 BIBPOLYTEC
TÉLÉCOPIEUR (FAX) : (514) 343-7121

CRT-635

A PRIMAL-DUAL ALGORITHM FOR SOLVING MONOTONE VARIATIONAL INEQUALITIES, AND ITS APPLICATION TO NETWORK EQUILIBRIUM

by

Patrice Marcotte*

* Centre de recherche sur les transports, Université de Montréal, and the Collège militaire royal de Saint-Jean, Richelieu, Quebec, CANADA.

Université de Montréal

Centre de recherche sur les transports - Publication #635

June 1989

Dépôt légal - 2e trimestre 1989

ABSTRACT

In this paper we propose a primal-dual implementation of Khobotov's algorithm for solving monotone variational inequalities over subsets of R^n. Besides its theoretical appeal, the algorithm lends itself nicely to a network implementation. An estimate of the local rate of convergence is given and numerical examples are presented.

Key words : networks, variational inequalities, mathematical programming.

- - - - - - -

RÉSUMÉ

Dans cet article nous proposons une implantation primale-duale de l'algorithme de Khobotov pour la résolution d'inéquations variationnelles dans R^n. En plus de propriétés théoriques intéressantes, l'algorithme se prête à une interprétation particulière lorsqu'implanté sur des réseaux. Nous donnons un estimé du taux de convergence local de la méthode ainsi que des résultats numériques préliminaires.

Mots clés : réseaux, inéquations variationnelles, programmation mathématique.

1. Khobotov's algorithm.

Let Φ be a closed convex subset of R^n and F a monotone, Lipschitz continuous mapping from Φ into R^n. A point x in Φ is a solution to the variational inequality problem $VIP(\Phi, F)$ if it satisfies the variational inequality:

$$\langle F(x), x - y \rangle \leq 0 \qquad \forall y \in \Phi. \tag{1}$$

The solution set S of $VIP(\Phi, F)$ is always convex. Assume that it is nonempty (basic notions required to understand this paper can be found in the work of Pang and Chan [6], for example).

A traditional approach for solving $VIP(\Phi, F)$ is through the projection algorithm:

$$x^{k+1} = \mathrm{Proj}_{\Phi}(x^k - \alpha F(x^k)) \tag{2}$$

where Proj_{Φ} denotes the projection operator on Φ in any suitable matrix norm and α is a constant whose value depends on both the Lipschitz and strong monotonicity constants of F. Those values might be difficult to estimate in practice. Khobotov's algorithm can be considered a two-step variant of the basic projection algorithm, and is defined by the relationship (see figure 1 for an illustration):

$$\begin{aligned} \bar{x}^k &= \mathrm{Proj}_{\Phi}(x^k - \alpha_k F(x^k)) \\ x^{k+1} &= \mathrm{Proj}_{\Phi}(x^k - \alpha_k F(\bar{x}^k)) \end{aligned} \tag{3}$$

where the relaxation parameter can easily be estimated *locally* and is *not* dependent on a strong monotonicity constant. The fundamental result is stated below and, for the sake of completeness, a proof is given in the appendix.

Theorem (Khobotov [3]). *Let Φ be a closed convex subset of R^n and F a monotone function from Φ into R^n. Let $x^0 \in \Phi$ and $\{x^k\}_{k \geq 1}$ be the sequence defined by (3). Then we have, for any nonnegative sequence $\{\alpha_k\}$ and $x^* \in S$:*

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \|x^k - \bar{x}^k\|^2 \left\{ 1 - \alpha_k^2 \frac{\|F(x^k) - F(\bar{x}^k)\|^2}{\|x^k - \bar{x}^k\|^2} \right\}. \tag{4}$$

□

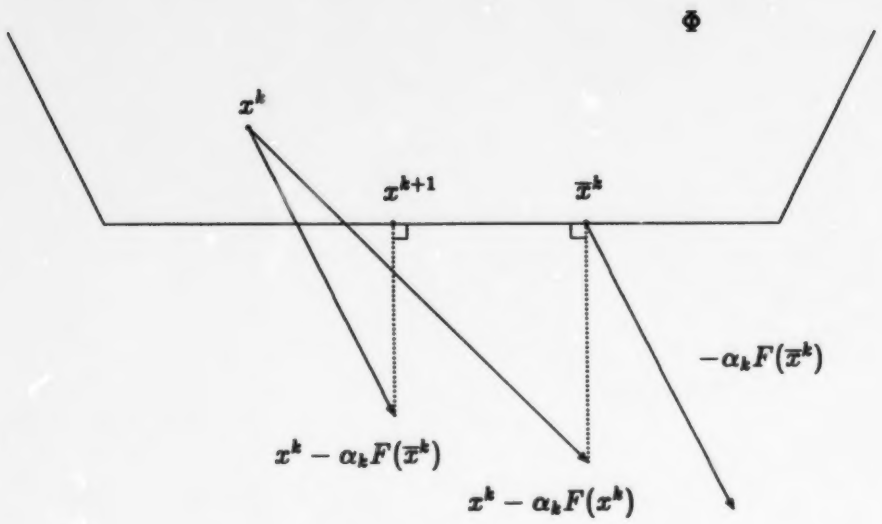

Figure 1. Geometry of Khobotov's algorithm.

If α_k is small enough, say strictly less than the Lipschitz constant L, then the inequality (4) shows that the distance to the set S strictly decreases at each iteration and we must have that $\lim_{k\to\infty} x^k = \overline{x}^k$ (otherwise the left-hand side of (4) would eventually become strictly negative). From the continuity of the projection operator there follows that any limiting point x^* of the sequence $\{x^k\}$ satisfies $x^* = \mathrm{Proj}_\Phi(x^* - \alpha_k F(x^*))$. If the sequence $\{\alpha_k\}$ stays bounded away from zero, x^* must be a solution to $VIP(\Phi, F)$. It is remarkable that the convergence proof is not based on a contraction theorem, but on the fact that the successive iterates get closer to the solution set by a positive quantity that can be bounded from below (see inequation (4)) by a *controllable* amount. At each iteration it is tempting to make the quantity:

$$h(\alpha) \stackrel{\text{def}}{=} \|x^k - \overline{x}^k\|^2 \left\{ 1 - \alpha_k^2 \frac{\|F(x^k) - F(\overline{x}^k)\|^2}{\|x^k - \overline{x}^k\|^2} \right\}$$

as large as possible. In that respect, an educated guess would consist in setting α_k to the value $(1/\sqrt{2})(\|x^k - \overline{x}^k\|/\|F(x^k) - F(\overline{x}^k)\|)$ that would be optimal, in the sense that the right-hand side of (4) is minimized, under the assumptions that (i) F is affine (ii) Φ is polyhedral (iii) x^k and $\overline{x}^k$ lie on the same face of Φ. In our numerical

affine (ii) Φ is polyhedral (iii) x^k and $\bar{x}^k$ lie on the same face of Φ. In our numerical tests this proved to be a reasonable strategy. The basic primal algorithm is outlined below:

Primal algorithm

1. Select parameters β $(0 < \beta < 1)$ and α $(\alpha > 0)$.
   Let $x^0 \in \Phi$ and set k to 1.
2. $\bar{x}^k = \text{Proj}_\Phi(x^k - \alpha F(x^k))$
3. if $\alpha^2(\|F(x^k) - F(\bar{x}^k)\|^2 / \|x^k - \bar{x}^k\|^2) \leq \beta^2$
   then $x^{k+1} = \text{Proj}_\Phi(x^k - \alpha F(\bar{x}^k))$, $k = k + 1$
   else Reduce the value of α, e.g.:

$$\alpha = \min\left\{\frac{\alpha}{2}, \frac{1}{\sqrt{2}} \frac{\|x^k - \bar{x}^k\|}{\|F(x^k) - F(\bar{x}^k\|}\right\}.$$

4. go to step 2.

□

The convergence of the sequence $\{x^k\}$ can be monitored through the upper bound on the decrease of the distance to the equilibrium set provided by inequation (4) or, if Φ is bounded, by evaluating the gap function (see Marcotte [5])

$$g(x^k) \stackrel{\text{def}}{=} \langle x^k - \bar{x}^k, F(x^k)\rangle$$

at the current iterate x^k. The gap function is nonnegative, and zero if and only if x^k is solution to the variational inequality.

We conclude this section with an analysis of the rate of convergence of the method when the function F is strongly monotone, i.e. there exists a constant l such that

$$\langle F(x) - F(y), x - y\rangle \geq l\|x - y\|^2 \qquad \text{for all } x, y \text{ in } \Phi.$$

Theorem. *Let f be strongly monotone on Φ with strong monotonicity constant l. Then the sequence x^k defined by the primal algorithm converges linearly to the unique equilibrium solution x^*. More precisely there exists an index K such that, for all $k \geq K$ we have:*

$$\|x^{k+1} - x^*\|^2 \leq \frac{\alpha_k^2 + \beta^2}{1 + \beta^2}\|x^k - x^*\|^2$$

where α is any positive constant satisfying $0 < \alpha < \liminf\{x^k\}$ and $\beta = \frac{1}{l}(L + \frac{1}{\alpha})$.

Proof. Existence and uniqueness of x^* is a consequence of the strong monotonicity assumption. This is a standard result (see Auslender [1] for example).

From (4) we have

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\le \|x^k - x^*\|^2 - \|x^k - \overline{x}^k\|^2\{1 - \alpha_k^2 \frac{\|F(x^k) - F(\overline{x}^k)\|^2}{\|x^k - \overline{x}^k\|^2}\} \\ &\le \|x^k - x^*\|^2 - \|x^k - \overline{x}^k\|^2\{1 - \alpha_k^2 L^2\} \\ &= \|x^k - x^*\|^2\{1 - (1 - \alpha_k^2 L^2)\frac{\|x^k - \overline{x}^k\|^2}{\|x^k - x^*\|^2}\}. \end{aligned} \tag{9}$$

Now, since $\overline{x}^k = \mathrm{Proj}_\Phi(x^k - \alpha_k F(x^k))$ we have:

$$\langle \overline{x}^k - y, \overline{x}^k - x^k + \alpha_k F(x^k)\rangle \le 0 \quad \forall y \in \Phi$$

or:

$$\langle \overline{x}^k - y, \overline{x}^k - x^k + \alpha_k[F(x^k) - F(\overline{x}^k)] + \alpha_k F(\overline{x}^k)\rangle \le 0 \quad \forall y \in \Phi.$$

Then:

$$\begin{aligned}\alpha_k\langle \overline{x}^k - y, F(\overline{x}^k)\rangle &\le -\langle \overline{x}^k - y, \overline{x}^k - x^k\rangle + \alpha_k\langle \overline{x}^k - y, F(\overline{x}^k) - F(x^k)\rangle \\ &\le \|\overline{x}^k - y\| \cdot \|\overline{x}^k - x^k\|(1 + \alpha_k L) \quad \forall y \in \Phi.\end{aligned}$$

Hence:

$$\langle \overline{x}^k - y, F(\overline{x}^k)\rangle \le (L + \frac{1}{\alpha_k})\|\overline{x}^k - y\| \cdot \|\overline{x}^k - x^k\| \quad \forall y \in \Phi.$$

Set $y = x^*$ in the preceding inequality to obtain:

$$\langle \overline{x}^k - x^*, F(\overline{x}^k)\rangle \le (L + \frac{1}{\alpha_k})\|\overline{x}^k - x^*\| \cdot \|\overline{x}^k - x^k\|.$$

We have:

$$\begin{aligned}\langle \overline{x}^k - x^*, F(\overline{x}^k) - F(x^*)\rangle &= \langle \overline{x}^k - x^*, F(\overline{x}^k)\rangle - \langle \overline{x}^k - x^*, F(x^*)\rangle \\ &\le \langle \overline{x}^k - x^*, F(\overline{x}^k)\rangle \\ &\le (L + \frac{1}{\alpha_k})\|\overline{x}^k - x^*\| \cdot \|\overline{x}^k - x^k\|.\end{aligned}$$

Therefore:

$$l\|\overline{x}^k - x^*\|^2 \le (L + \frac{1}{\alpha_k})\|\overline{x}^k - x^*\| \cdot \|\overline{x}^k - x^k\|$$

and:

$$\|\bar{x}^k - x^*\|^2 \leq \frac{1}{l}(L + \frac{1}{\alpha})\|\bar{x}^k - x^k\| = \beta\|\bar{x}^k - x^k\|$$

whenever $k \geq K$. Now:

$$\begin{aligned}\|x^k - x^*\|^2 &\leq \|x^k - \bar{x}^k\|^2 + \|\bar{x}^k - x^*\|^2 \\ &\leq (1+\beta^2)\|x^k - \bar{x}^k\|^2\end{aligned}$$

which implies

$$\frac{\|x^k - \bar{x}^k\|^2}{\|x^k - x^*\|^2} \geq \frac{1}{1+\beta^2}. \tag{10}$$

Using (10) in (9) there comes:

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2\{1 - \frac{1-\alpha_k^2 L^2}{1+\beta^2}\} \\ &= \frac{\beta^2 + \alpha_k^2 L^2}{\beta^2 + 1}\|x^k - x^*\|^2.\end{aligned}$$

□

Remark. The preceding estimate of the local convergence rate is in many situations overly pessimistic. For instance, in the simple case where $F(x) = \|x\|^2$ the preceding estimate would be close to 1 for α close to 1, whereas setting α to 1 yields the optimal solution in a single iteration.

2. A primal-dual algorithm.

In general it is computationally burdensome to perform a projection onto an arbitrary convex set. In the following we assume that the set Φ is polyhedral:

$$\Phi = \{Bx = b, x \geq 0\}$$

where B is an $m \times n$ matrix. The analysis carries through if Φ can be expressed by a set of nonlinear equations, and a certain constraint qualification holds. Then $VIP(\Phi, F)$ can, using linear programming duality, be put in the form:

$$\begin{aligned} F(x) - B^t\lambda &\geq 0 \\ \langle x, F(x) - B^t\lambda\rangle &= 0 \\ Bx &= b \\ x &\geq 0 \\ \lambda &\quad \text{unconstrained}\end{aligned} \tag{11}$$

which is easily seen to be equivalent to the monotone variational inequality $VI(\Psi, G)$ where $\Psi = \{x \geq 0, \lambda \text{ unconstrained}\}$ and $G\begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} F(x) - B^t\lambda \\ Bx - b \end{pmatrix}$. In this primal-dual setting, Khobotov's algorithm reduces to:

$$\begin{aligned} \overline{x}_i^k &= \max\{0, x_i^k - \alpha_k[F_i(x^k) - (B^t\lambda^k)_i]\} \\ \overline{\lambda}_j^k &= \lambda_j^k - \alpha_k(Bx^k - b)_j \end{aligned} \tag{12a}$$

and

$$\begin{aligned} x_i^{k+1} &= \max\{0, x_i^k - \alpha_k[F_i(\overline{x}^k) - (B^t\overline{\lambda}^k)_i]\} \\ \lambda_j^{k+1} &= \lambda_j^k - \alpha_k(B\overline{x}^k - b)_j. \end{aligned} \tag{12b}$$

Remark. It can be shown, using a technique a proof similar to that used by Marcotte and Dussault [4], that the dual vectors λ^k are bounded if the set Φ is bounded (the set Ψ need *not* be bounded).

3. An application to network equilibrium.

In this section we consider single commodity network equilibrium problems (traffic assignment, spatial equilibrium, etc.) that can be formulated as a variational inequality. In this case, the matrix $B = (B_{ij})$ can be identified with the node-arc incidence matrix of the network. If there exists a directed arc $a = (i, j)$ between nodes i and j then $B_{ji} = -1$ and $B_{ij} = +1$. The right-hand side (production) vector b is positive at origin (production) nodes, negative at destination (demand) nodes and zero at transshipment (intermediate) nodes. Equations (12a) and (12b) can then be written as:

$$\text{arc } (i,j): \qquad \overline{x}_{ij}^k = max\{0, x_{ij}^k - \alpha_k[F_{ij}(x^k) + \lambda_i^k - \lambda_j^k]\} \tag{13a}$$

$$\text{node } i: \qquad \overline{\lambda}_i^{k+1} = \lambda_i^k - \alpha_k[\sum_j x_{ij}^k - \sum_j x_{ji}^k - b_i]. \tag{13b}$$

$$\text{arc } (i,j): \qquad x_{ij}^{k+1} = \max\{0, x_{ij}^k - \alpha_k[F_{ij}(\overline{x}^k) + \overline{\lambda}_i^k - \overline{\lambda}_j^k]\} \tag{14a}$$

$$\text{node } i: \qquad \lambda_i^{k+1} = \lambda_i^k - \alpha_k[-\sum_j \overline{x}_{ij}^k + \sum_j \overline{x}_{ji}^k - b_i]. \tag{14b}$$

The quantity $r_{ij}^k \stackrel{\text{def}}{=} F_{ij}(x^k) + \lambda_i^k - \lambda_j^k$ can be interpreted as the "reduced" cost on arc (i, j) and $\delta_i \stackrel{\text{def}}{=} \sum_j x_{ij}^k - \sum_j x_{ij}^k - b_i$ as the divergence at node i. Equations (14a) and

(14b) can be given an intuitive network interpretation. If $\overline{x}$ and $\overline{\lambda}$ are close to x and λ respectively, then it seems reasonable to increase λ_i whenever divergence is negative, in order to reduce the amount of flow coming *into* node i, through the increase of the reduced costs of all arcs having i as their tail. If divergence is positive then the variable λ_i is decreased. Also, flow is increased (respectively decreased) on arcs with negative (respectively positive) reduced costs, as in the out-of-kilter algorithm.

4. Numerical examples.

Braess network. Consider the Braess paradox network with the separable linear cost function illustrated in figure 2. Demand from node 1 to node 4 is equal to 6 units of flow, and the delay functions are given by the expressions:

$$F_{12}(x) = 10\,x_{12}$$
$$F_{34}(x) = 10\,x_{34}$$
$$F_{23}(x) = 10 + x_{23}$$
$$F_{13}(x) = 50 + x_{13}$$
$$F_{24}(x) = 50 + x_{24}.$$

The ordering of the arcs is: (1,2), (1,3), (2,3), (2,4), (3,4).

Starting with the initial solution

$$x = (6,0,6,0,6)$$
$$\lambda = (0,0,0,0)$$

we obtain the sequence of iterates given in table 1. In this example the matrix B and the vector b have been premultiplied by the constant 5. This is equivalent to a scaling of the cost function F. This multiplicative constant seems to affect the algorithm to a great extent. Future work will be devoted to determining a good value for this parameter.

The equilibrium solution for this problem is:

$$x^* = (4,2,2,2,4)$$
$$\lambda^* = (0,40,42,92)$$

if one always sets to zero the dual variable associated with the origin node 1.

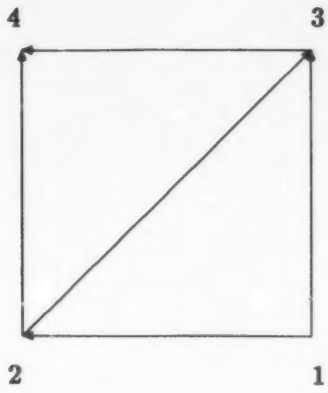

Figure 2. The Braess paradox network.

At the last iteration. feasibility can be restored by projecting the final flow vector on Φ.

Remark 1. The work per iteration involves $m + n$ variable adjustments and m function evaluations and is therefore $O(m)$ for a single commodity problem.

Remark 2. It can be observed that the distance to equilibrium $\|x - x^*\|$ does not decrease monotonically. This is not in contradiction with the theory. In our primal-dual setting the distance to the equilibrium set involves both primal and dual variables and is indeed monotonically decreasing, although this measure is not reported in table 1.

Second network example. The second example corresponds to the network illustrated in figure 3. The data for this problem is given in table 2. The congestion function is arc-separable and assumes the general (BPR) form:

$$F_{ij}(x) = \alpha_{ij} + \beta_{ij}(x_{ij}/y_{ij})^4.$$

Numerical results are to be found in table 3.

Sioux Falls network. The third network corresponds to the Sioux Falls network (see Abdulaal and Leblanc [7]). It has 24 nodes and 76 arcs and, in our case, 4

k							$\|x-x^*\|$ Khobo.	$\|x-x^*\|$ FW
1	x:	3.46	0.00	5.06	0.00	3.46	4.24	3.38
	λ:	0.00	4.72	0.73	5.45			
2	x:	3.14	0.00	3.74	0.00	3.14	3.54	1.00
	λ:	0.00	14.28	3.76	18.06			
3	x:	2.99	0.00	2.00	0.00	2.99	3.17	.062
	λ:	0.00	21.45	10.25	31.70			
4	x:	3.08	0.00	0.42	0.00	3.08	3.49	.042
	λ:	0.00	24.89	22.58	47.46			
5	x:	3.07	0.00	0.37	0.00	3.07	3.52	.030
	λ:	0.00	26.61	34.44	61.05			
10	x:	4.30	1.79	2.55	1.79	4.30	0.07	.068
	λ:	0.00	44.20	51.03	95.23			
15	x:	4.11	2.05	2.10	2.05	4.11	0.02	.017
	λ:	0.00	40.84	51.67	92.51			
20	x:	4.01	2.01	2.03	2.01	4.01	0.04	.005
	λ:	0.00	40.08	51.99	92.09			
30	x:	4.00	2.00	2.00	2.00	4.00	.002	.0004
	λ:	0.00	40.00	52.01	92.01			

Table 1. Numerical results for the Braess network example.

origin-destination pairs. It is therefore a multicommodity problem. For this example we used the gap function

$$g(x) = \max_{y \in \Phi} \langle x - y, F(x) \rangle$$

as a measure of equilibrium. However, since this quantity only makes sense when x is close to being primal-feasible, we indicate as well, in table 4, the infeasibility associated with the current iterate x^k. On all three test problems the convergence

pattern was similar. Progress was slow in the first iterations, but fast once a good approximation was obtained. This was especially true of the dual variables, converging very slowly to their optimal values. Dual update methods are presently being considered. Also, the good behaviour of the Frank-Wolfe algorithm could be used to "warm-up" the algorithm, even though this would be heuristic in the asymmetric case. Let us notice that on the first two test problems, Frank-Wolfe's good performance was partially due to the fact that the first iterate lied in the optimal face which indeed corresponds to the whole set Φ in these cases, thus ensuring linear local convergence. In general, Frank-Wolfe cannot be expected to converge at a better than sublinear (harmonic) rate.

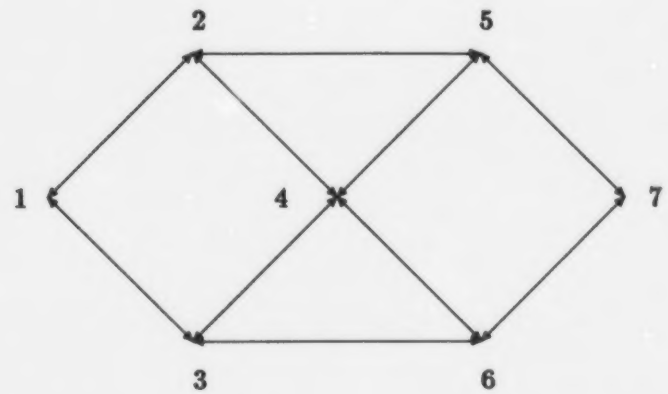

Figure 3. Second test network.

Conclusion.

In this paper we presented a projection-based primal-dual algorithm for solving monotone variational inequalities, and its implementation for network equilibrium

arc	arc	α	β	y
1	(1,2)	3	3	75
2	(1,3)	3	3	75
3	(2,3)	3	3	75
6	(3,1)	3	3	75
18	(6,7)	3	3	75
19	(7,5)	3	3	75
4	(2,4)	4	4	25
7	(3,4)	4	4	25
9	(4,2)	4	4	25
11	(4,5)	4	4	25
12	(4,6)	4	4	25
14	(5,4)	4	4	25
17	(6,4)	4	4	25
8	(3,6)	8	8	50
13	(5,2)	8	8	50
16	(6,3)	8	8	50
5	(2,5)	8	8	20
10	(4,3)	4	4	5
15	(5,7)	3	1	95

Demand

$1 \longrightarrow 4 \quad : \quad 20$

$7 \longrightarrow 1 \quad : \quad 60$

$1 \longrightarrow 7 \quad : \quad 80$

$7 \longrightarrow 4 \quad : \quad 30$

Table 2. Data for second test network.

problems. The extension to multicommodity problems is straightforward, since the projection separates by commodity if one is willing to store flows by commodity. Although the algorithm might not be competitive with other codes for solving sym-

k	$\|x^k - x^*\|$ Khobotov	$\|x^k - x^*\|$ FW
10	136.6	13.4
50	45.3	12.1
100	7.3	10.7
150	1.8	9.8
200	.36	9.0
250	.07	8.5
300	.01	8.0

Table 3. Numerical results, second example.

k	$g(x^k)$	$\|Bx^k - b\|$
10	5649	180.4
50	2155	127.2
100	13.6	.38
150	3.73	.096
200	.778	.0146
250	.156	.0029
300	.043	.0008
600	2×10^{-6}	4×10^{-8}

Table 4. Sioux Falls results.

metric problems, it possesses several theoretical and practical advantages over current algorithms:

1. global convergence for monotone problems is assured
2. the relaxation parameter α_k is easily adjusted at each iteration, upwards or downwards
3. the algorithm takes advantage of the network structure
4. the amount of memory space required by the algorithm can be estimated *a priori*

5. the method does not require derivatives
6. there is flexibility for scaling the cost mapping F, for instance by using different norms
7. the algorithm is easily parallelized
8. no linesearch is required.

Let us also note that the algorithm performed well on non-network variational inequality problems, even in situations where the classical projection algorithm would fail, for instance for affine, skew-symmetric problems.

Recently, Bertsekas [1] [2] obtained impressive computational results using a dual ascent method closely related to the out-of-kilter algorithm for solving linear as well as convex, separable flow problems. His approach is not readily applicable to non symmetric problems. However the price and flow adjustment steps bear some resemblance to our primal-dual adjustments (equations (7a) and (7b)). A drawback of our method is that nodes *cannot* be treated in an arbitrary order, as in Bertsekas [2]. In particular, there is no obvious equivalent for the dual function in variational inequalities. Future work will concentrate on increasing the flexibility of node and arc treatment, in order to be able to exploit the power of network data structures, always keeping in mind that (costly) function evaluations have to be kept to a minimum.

REFERENCES

[1] Bertsekas, D. P., Hossein, P. A. and Tseng, P. (1987), "Relaxation methods for networks flow problems with convex arc costs", SIAM Journal on Control and Optimization 25,2 pp 1219–1243.

[2] Bertsekas, D. P. and Tseng, P. (1988), "Relaxation methods for minimum cost ordinary and generalized network flow problems", Operations Research 36,1 pp 93–114.

[3] Khobotov, E. N. (1987), "A modification of the extragradient method for the solution of variational inequalities and some optimization problems", Zh. Vychisl. Mat i Mat. Fiz. 27, 10 pp 1462–1473.

[4] Marcotte, P. and Dussault, J.-P. (1985), "A modified Newton method for solving variational inequalities", Proceedings of the 24^{th} IEEE Conference on Decision and Control, Fort Lauderdale, pp 1433–1436.

[5] Marcotte, P. (1985), "A new algorithm for solving variational inequalities, with application to the traffic assignment problem", Mathematical Programming 33, pp 339–351.

[6] Pang, J. S. and Chan, D. (1984), "Iterative methods for variational and complementarity problems", Mathematical Programming 24, pp 284–313.

[7] Abdulaal, M. And LeBlanc, L. J. (1979), "Continuous equilibrium design models", Transportation Research 13B pp 19–32.

APPENDIX

A STREAMLINED PROOF OF KHOBOTOV'S THEOREM

For any two vectors u and v ($v \in \Phi$) we have:

$$\|u - v\|^2 \geq \|u + \text{Proj}_\Phi(u)\|^2 - \|v - \text{Proj}_\Phi(u)\|^2. \tag{15}$$

Setting $u = x^k - \alpha_k F(\overline{x}^k)$, $v = x^*$ ($x^* \in S$), $x^{k+1} = \text{Proj}_\Phi(u)$ in (15) and rearranging terms, one obtains the inequality:

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\leq \|x^k - \alpha_k F(\overline{x}^k) - x^*\|^2 - \|x^k - \alpha_k F(\overline{x}^k) - x^{k+1}\|^2 \\ &= \|x^k - x^*\|^2 - \|x^k - x^{k+1}\|^2 + 2\alpha_k \langle x^* - x^{k+1}, F(\overline{x}^k)\rangle. \end{aligned} \tag{16}$$

We have:

$$\begin{aligned}\langle x^* - x^{k+1}, F(\overline{x}^k)\rangle &= \langle x^* - \overline{x}^k, F(\overline{x}^k)\rangle + \langle \overline{x}^k - x^{k+1}, F(\overline{x}^k)\rangle \\ &\leq \langle \overline{x}^k - x^{k+1}, F(\overline{x}^k)\rangle, \end{aligned} \tag{17}$$

since $x^* \in S$ and F is monotone.

There follows, using (17) in (16):

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - \|x^k - x^{k+1}\|^2 + 2\alpha_k \langle \overline{x}^k - x^{k+1}, F(\overline{x}^k)\rangle \\ &= \|x^k - x^*\|^2 - \|x^k - \overline{x}^k\|^2 - \|\overline{x}^k - x^{k+1}\|^2 - 2\langle x^k - \overline{x}^k, \overline{x}^k - x^{k+1}\rangle \\ &\quad + 2\alpha_k \langle \overline{x}^k - x^{k+1}, F(\overline{x}^k)\rangle \\ &= \|x^k - x^*\|^2 - \|x^k - \overline{x}^k\|^2 - \|\overline{x}^k - x^{k+1}\|^2 \\ &\quad + 2\langle x^k - \alpha_k F(\overline{x}^k) - \overline{x}^k, x^{k+1} - \overline{x}^k\rangle \\ &= \|x^k - x^*\|^2 - \|x^k - \overline{x}^k\|^2 - \|\overline{x}^k - x^{k+1}\|^2 \\ &\quad + 2\langle x^k - \alpha_k F(x^k) - \overline{x}^k, x^{k+1} - \overline{x}^k\rangle + 2\alpha_k \langle F(x^k) - F(\overline{x}^k), x^{k+1} - \overline{x}^k\rangle \\ &\leq \|x^k - x^*\|^2 - \|x^k - \overline{x}^k\|^2 - \|\overline{x}^k - x^{k+1}\|^2 \\ &\quad + 2\alpha_k \|F(x^k) - F(\overline{x}^k)\| \cdot \|x^{k+1} - \overline{x}^k\| \end{aligned} \tag{18}$$

since $\langle \text{Proj}_\Phi(u) - u, \text{Proj}_\Phi(u) - v\rangle \leq 0$ if $v \in \Phi$.

Using the inequality

$$\|x^{k+1} - \overline{x}^k\|^2 + \alpha_k^2 \|F(x^k) - F(\overline{x}^k)\|^2 \geq 2\alpha_k \|F(x^k) - F(\overline{x}^k)\| \cdot \|x^{k+1} - \overline{x}^k\|$$

in (18) there comes:

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 \leq & \|x^k - x^*\|^2 - \|x^k - \overline{x}^k\|^2 - \|\overline{x}^k - x^{k+1}\|^2 \\
& + \|x^{k+1} - \overline{x}^k\|^2 + \alpha_k^2 \|F(x^k) - F(\overline{x}^k)\|^2 \\
\leq & \|x^k - x^*\|^2 - \|x^k - \overline{x}^k\|^2 \left\{ 1 - \alpha_k^2 \frac{\|F(x^k) - F(\overline{x}^k)\|^2}{\|x^{k+1} - \overline{x}^k\|^2} \right\}.
\end{aligned}$$